

Capstone Project - The Battle of Neighborhoods (Week 4)

Week 4. Problem Description, Approach and the Data used

Discussion of the background

A shopping mall is a modern form of shopping precinct or shopping center in which one or more buildings form a complex of shops with interconnecting walkways, usually indoors. They combine a number of events under one roof, namely, essentials and cloth shopping, restaurants and dining, watching movies and a number of other activities depending on the size and kind of mall such as playing games. Shopping malls are a one-stop destination for all types of recreation and shopping activities. For individual shops, malls provide a great distribution channel to market their products and services through the abundant crowd that flows in and out everyday. Property developers take advantage of this trend to build more shopping malls to cater to the ever-increasing demand of the public. As a result, Bangalore - the IT capital of India has experienced rapid urbanization with a number of newer malls being opened every year. As with any business decision, opening a new shopping mall requires a multi-faceted consideration of the local geography and crowds surrounding those areas. In particular, the location of the shopping mall is one of the most important decisions that determines the footfall and invariably the success or failure of a mall.

Description of the problem

The primary objective of this capstone project is to understand the local geography of the metropolitan city of Bengaluru (old name Bangalore) and come up with a suitable selection of the best locations to construct a new shopping mall. Using hitherto-learned Data Science from the previous weeks and Machine-Learning techniques like k-Means Clustering, this project aims to answer the business question: In which areas should a property developer invest in to build a new mall in order to maximise his/her profit in Bangalore, India?

Who will this benefit?

This Data Science project is aimed at providing an answer or a starting point to property developers and investors who're on the look out to construct new shopping malls in the IT capital of India - Bengaluru. This project is timely as the city is facing rapid urbanization with an influx of people from all across the country. This necessitates greater infrastructure to cater to the diverse needs of the people and ensuring that one particular neighborhood doesn't suffer from oversupply while another suffers from undersupply is of paramount importance. Data from the Anarock Property Consultants report showed that an additional 65 million sq ft will be added to existing mall space by 2022. The local newspaper Deccan herald also highlighted the surge in the number of malls the country will face in the coming years.

Data

Description of the Data

To solve the defined problem above, we make use of the following data:

- List of neighbourhoods in Bangalore from the Wikipedia page(https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore)
- Latitude and longitude coordinates of those neighborhoods from the *Geocoder* python library. This is required in order to plot the map using and also to get the venue data.
- Using the Foursquare API to get venue data, particularly data related to shopping malls. We will use this data to perform clustering in the neighbourhoods.

Sources of the data and how they're used to solve the problem

1. Wikipedia (https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore)

The listed webpage enlist all the neighbourhoods in Bangalore, with a total of 72 neighbourhoods in the *Bangalore Urban* District. We use web scraping techniques to extract the data from the Wikipedia page, as done previously in clustering neighborhoods in Toronto with the help of Python requests and BeautifulSoup packages.

2. Geocoder

Next, we get the geographical coordinates of the neighbourhoods using the *Geocoder* package which will give us the latitude and longitude coordinates of each neighborhood.

3. Foursquare API

We then use Foursquare API to get the venue data of those neighbourhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API provides various categories of venue data, out of which we are particularly interested in the *Shopping Mall* category in order to help us solve the business problem defined.

This is a project that makes use of many Data Science skills, from web scraping Wikipedia pages to working with the Foursquare API, data cleaning, data wrangling and Machine Learning (k-Means Clustering) and data map visualization (Folium). The detailed steps of the approach are written in the final report and the Final Notebook of the project containing the code.